HW 01: Data Entry

Recording data to maintain your sanity.

Assignment Overview

Although you will be working with previously collected data, it is important to understand what data looks like as well as how it is coded and entered into a spreadsheet or dataset for analysis. This can help you identify and avoid problems later when reading data into an analysis software program. For example if you mix letters and numbers in the same cell, the variable will be treated as character not numeric.

Instructions

Preparation: Read Tidy data principles by Wickham, Vaughan, and Girlich.

There are three pieces to this assignment.

1. Entering raw data into a spreadsheet.
2. Creating a codebook
3. Importing the data into R

Using the PDF copies of medical records for 5 patients seeking treatment in a hospital emergency room you are going to do data entry and create a codebook.

Where to do the work

You will enter data and create your codebook directly in Google Sheets

- Start a new Google spreadsheet in the **01 Data Entry** folder in our shared Google Drive.
- Name this file `medrecords_userid` where *userid* is your chico state user id.
- Create and name the following worksheets: `data`, `codebook`, `import`.

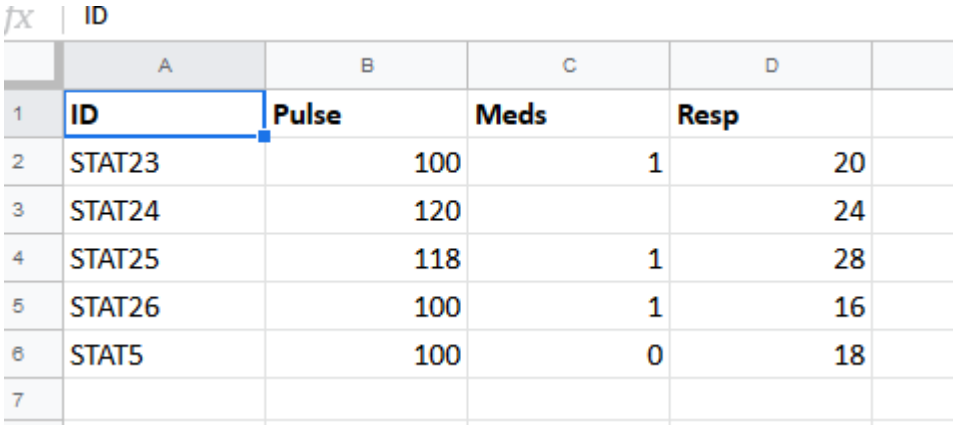
Submission Instructions

Download your google workbook (all sheets) as an Excel file (.xlsx or .xls) and upload to Canvas by the due date.

Assignment

1. Data Entry

1. Select 4 variables recorded on the medical forms
    - one should be a unique identifier, at least one should be a quantitative variable and at least one should be a categorical variable. (Ref PMA6 Ch2)
2. Select a brief name for each variable - write this in the first row
    - Use good variable naming conventions:
        - short
        - no special characters
        - no spaces
        - doesn't start with a number
3. Determine what range of values is needed for recording each variable
4. Enter the data for each patient, one patient per row.
5. If data is missing for a particular value, leave the cell blank.



Recall the *tidy data principles* state to put one observation per row, and one variable (characteristic) per column.

2. Codebook Creation

In a separate worksheet list the variable names, labels, data types, and response code or ranges in separate columns (4 columns total).

An example of what this should look like is below. (With the exception of the red error)

Variable	Label	Type	Response codes
ID	identiication	unique identifier	
Gender	Gender	categorical	0= male 1= female
BP	blood pressure	quantitatve	what are plausible ranges? is 0 ok?
CV	review of cardiovascular system	categorical	0 = yes 1 = no 2 = unknown

3. Data Import

1. Export your file to your hard drive as a Comma Separated Value (`*.csv`).
    - If it asks you, only save the`data` worksheet.
2. Import this data into R.
    - Code is fine if you already know how. Point and click is also fine for now.
    - Point and click instructions
    - Code examples from R for data science and Statology
3. Note and record any problems that you noticed and/or had to fix at the bottom of your `codebook` worksheet.
    - Does your data file look like your spreadsheet?
    - Did you have to specify missing values in any specific way?
4. Show that it worked by taking a screenshot of BOTH the code, and the view of the data.
    - Paste these images into your `import` tab.

Examples of import code

```
medrecords_iainudidn <- read.csv("~/Desktop/MATH615/assignments/project1/data/medrecords_iainudidn.csv")
```

Figure 1: Example R code